

# Training Set

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

# Training Set

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as  $\mathcal{D}$ , containing:

1. Feature matrix  $\mathbf{X} \in \mathbb{R}^{n \times d}$

# Training Set

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as  $\mathcal{D}$ , containing:

1. Feature matrix  $\mathbf{X} \in \mathbb{R}^{n \times d}$
2. Output vector  $\mathbf{y} \in \mathbb{R}^n$

## Entropy calculated

We have 14 examples in  $S$ : 5 No, 9 Yes

$$\begin{aligned}\text{Entropy}(S) &= -p_{\text{No}} \log_2 p_{\text{No}} - p_{\text{Yes}} \log_2 p_{\text{Yes}} \\ &= -\frac{5}{14} \log_2 \left( \frac{5}{14} \right) - \frac{9}{14} \log_2 \left( \frac{9}{14} \right) = 0.940\end{aligned}$$

## Algorithm: Finding the Optimal Split

1. Sort all data points  $(x_i, y_i)$  in increasing order of  $x_i$ .

## Algorithm: Finding the Optimal Split

1. Sort all data points  $(x_i, y_i)$  in increasing order of  $x_i$ .
2. Evaluate the loss function for all candidate splits:

$$s = \frac{x_i + x_{i+1}}{2} \text{ for } i = 1, 2, \dots, n - 1$$

3. Select the split  $s^*$  that minimizes the loss function.